Linguaggi di Programmazione III appello – Soluzione traccia

1) Si individui quale delle seguenti affermazioni è corretta.
   Scegli un'alternativa:

   - ○ a. La tabella dei simboli è creata dal parser.
   - ◉ b. La tabella dei simboli è creata dallo scanner.
   - ○ c. La tabella dei simboli è creata dall’analizzatore semantico.
   - ○ d. La tabella dei simboli è creata dal generatore di codice.

(PUNTI 2)

2) Sia dato il seguente linguaggio L di alfabeto $X = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\}$:

   $$L = \{w \in X^* \mid w \neq \alpha\, x_1\, x_2\, \beta, \alpha \in X^*, \beta \in X^*, x_1 \in X, x_2 \in X, \}$$

   ove

   x_1 = primo carattere del proprio cognome

   x_2 = primo carattere del proprio nome

   (ad esempio, se il proprio cognome è “*rossi*” ed il proprio nome è “*mario*” allora $x_1 = r$, $x_2 = m$)

   Costruire, commentando opportunamente, un automa a stati finiti deterministico M con funzione di transizione totale che riconosce il linguaggio L (ossia, tale che T(M)= L).

(PUNTI 4)

Soluzione per cognome “*rossi*” e nome “*mario*” per cui $x_1 = r$, $x_2 = m$.

$$M = (Q, \delta, q_0, F)$$

$Q = \{q_0, q_1, q_2\}$

$\delta: Q \times X \to Q$

$\delta(q_0, r) = q_1$, $\delta(q_0, x) = q_0$ per ogni $x \in X - \{r\}$,

$\delta(q_1, m) = q_2$, $\delta(q_1, r) = q_1$, $\delta(q_1, x) = q_0$ per ogni $x \in X - \{r, m\}$

$\delta(q_2, x) = q_2$ per ogni $x \in X$

$F = \{q_0, q_1\}$

Determinare una grammatica lineare destra *G* corretta per *L* (ossia, tale che *L*(*G*)=*L*).

(PUNTI 2)

Applico l'Algoritmo 7.2.

G = (X, V, S, P)

V = $\{q_0, q_1, q_2\}$

S= q_0

$q_0 \to \lambda \mid r\, q_1 \mid r \mid x\, q_0 \mid x$ per ogni $x \in X - \{r\}$

$q_1 \to m\, q_2 \mid r\, q_1 \mid r \mid x\, q_0 \mid x$ per ogni $x \in X - \{r, m\}$

$q_2 \to x\, q_2$ per ogni $x \in X$

q_2 è un nonterminale inutile per cui G si può semplificare come segue:

V = $\{q_0, q_1\}$

S= q_0

$q_0 \to \lambda \mid r\, q_1 \mid r \mid x\, q_0 \mid x$ per ogni $x \in X - \{r\}$

$q_1 \to r\, q_1 \mid r \mid y\, q_0 \mid y$ per ogni $y \in X - \{r, m\}$

Determinare un'espressione regolare *R* che denota *L* (ossia, tale che *S*(*R*)=*L*).

Impostiamo il sistema di equazioni derivante dalla grammatica:

$q_0 = \lambda + r\, q_1 + r + R_1\, q_0 + R_1$ con R_1=(a + b + c + d + e + f + g + h + i + j + k + l + m + n + o + p + q + s + t + u + v + w + x + y + z)

$q_1 = r\, q_1 + r + R_2\, q_0 + R_2$ con R_2=(a + b + c + d + e + f + g + h + i + j + k + l + n + o + p + q + s + t + u + v + w + x + y + z)

Dalla seconda equazione, applicando la proprietà (20) sull'equivalenza delle espressioni regolari si ottiene:

$q_1 = r^* (r + R_2\, q_0 + R_2)$

Sostituendo q_1 nella prima equazione si ottiene:

$q_0 = \lambda + r\, r^* (r + R_2\, q_0 + R_2) + r + R_1\, q_0 + R_1 =$

$= \lambda + r\, r^* r + r\, r^* R_2\, q_0 + r\, r^* R_2 + r + R_1\, q_0 + R_1 = = (r\, r^* R_2 + R_1)\, q_0 + \lambda + r\, r^* r + r\, r^* R_2 + r + R_1$

Applicando la proprietà (20) sull'equivalenza delle espressioni regolari si ottiene l'espressione regolare richiesta:

$q_0 = (r\, r^*\, R_2 + R_1)^*\, (\lambda + r\, r^* r + r\, r^*\, R_2 + r + R_1)$

(PUNTI 2)

3) Sia dato il seguente linguaggio L di alfabeto $X = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$:

$$L = \{w \in X^* \mid w = x_1^i\, x_2^k\, x_3^j,\ k \geq i+j,\ i > 0,\ j > 0,\ x_1 \in X,\ x_2 \in X,\ x_3 \in X\}$$

ove

x_1 = primo carattere del proprio nome

x_2 = secondo carattere del proprio nome

x_1 = terzo carattere del proprio nome

(ad esempio, se il proprio nome è "*mario*" allora $x_1 = m$, $x_2 = a$, $x_3 = r$)

Stabilire in quale classe (*indicare la classe più specifica*) della gerarchia di Chomsky ricade L.

Giustificare formalmente la risposta.

L è un linguaggio di tipo 2.

Soluzione per nome "*mario*" per cui $x_1 = m$, $x_2 = a$, $x_3 = r$.

Giustifichiamo formalmente la risposta.

$$L = \{w \in X^* \mid w = m^i\, a^k\, r^j,\ k \geq i+j,\ i > 0,\ j > 0,\ x_1 \in X,\ x_2 \in X,\ x_3 \in X\}$$

Osserviamo che $L = L_1\, L_2\, L_3$.

ove

$L_1 = \{w \in X^* \mid w = m^i a^i,\ i > 0\}$

$L_2 = \{w \in X^* \mid w = a^i,\ i \geq 0\} = \{a\}^*$

$L_3 = \{w \in X^* \mid w = w = a^i r^j,\ i > 0\}$

L_1 è un linguaggio di tipo 2 in quanto G_1: S_1->ma | mS_1a è una grammatica di tipo 2 corretta per L_1 ($L(G_1)= L_1$).

L_2 è un linguaggio di tipo 3 in quanto $L_2 = (L'_2)^*$ ove ed L'_2 è di tipo 3 in quanto G'_2: S_2->a è una grammatica di tipo 3 corretta per L'_2 ($L(G'_2)= L'_2$) e la classe dei linguaggi di tipo 3 è chiusa rispetto all'operazione di iterazione.

L_3 è un linguaggio di tipo 2 in quanto G_3: S_3->ar | aS_3r è una grammatica di tipo 2 corretta per L_3 ($L(G_3)= L_3$).

L è un linguaggio di tipo 2 in quanto concatenazione di 3 linguaggi di tipo 2 (L_2 è un linguaggio di tipo 3 e dunque anche di tipo 2 per il teorema della gerarchia di Chomsky) e la classe dei linguaggi di tipo 2 è chiusa rispetto all'operazione di concatenazione.

Al fine di dimostrare che L non è un linguaggio di tipo 3, applichiamo il Pumping lemma per i linguaggi regolari.

Supponiamo per assurdo che esista un FSA $M = (Q, \delta, q_0, F)$ di alfabeto X e con $p=|Q|$ stati che riconosce L, ossia tale che $T(M)= L$.

Consideriamo la parola di L:

$$z = m^p a^{2p} r^p$$

Poiché $z \in L = T(M)$ si ha che $\delta^*(q_0, z) \in F$ per definizione di parola accettata da un FSA.

Dunque possiamo pensare che l'automa a stati finiti M per riconoscere il prefisso m^p della parola z abbia il seguente comportamento:

$\delta^*(q_0, m)=q_1$

$\delta^*(q_0, m^2)=q_2$

$\delta^*(q_0, m^3)=q_3$

…

$\delta^*(q_0, m^p)=q_p$

I $p+1$ stati $q_0, q_1,\dots q_p$ non possono essere tutti distinti poiché M ha $p=|Q|$ stati, dunque almeno uno stato tra $q_0, q_1,\dots q_p$ deve comparire 2 volte. Supponiamo che

$$q_i=q_j \text{ con } i<j \text{ ed } i,j \in \{0, 1, 2,\dots p\}$$

si ha: $\delta^*(q_0, m^i)=q_i=q_j=\delta^*(q_0, m^j)$

ossia esiste un ciclo di lunghezza j-i sullo stato q_i=q_j.

Possiamo dunque scrivere z come segue:

$$z = uvw$$

ove:

$u = m^i$

$v = m^{j-i}$

$w = m^{p-j}a^{2p}r^p$

Poiché $z \in T(M)$, per effetto dell'ingresso di $z=uvw$, l'automa M si porta per definizione in uno stato finale ($\delta^*(q_0, uvw) \in F$).

Ma è immediato osservare che tale stato è lo stesso in cui M si porta per effetto dell'ingresso della parola uv^2w dunque $\delta^*(q_0, uv^2w) \in F$ e uv^2w è accettata da M: $uv^2w \in T(M)$.

Ma $uv^2w = m^i\, m^{2(j-i)}\, m^{p-j}a^{2p}r^p = m^{p+j-i}a^{2p}r^p \notin L$.

Dunque $uv^2w \in T(M)$ e $uv^2w \notin L$ e siamo pervenuti ad una contraddizione.

La contraddizione deriva dall'aver assunto che che esista un FSA $M = (Q, \delta, q_0, F)$ di alfabeto X e con p=|Q| stati che riconosce L, ossia tale che T(M)= L.

Dunque per il teorema di Kleene possiamo concludere che L non è un linguaggio di tipo 3.

(PUNTI 6)

4) Si individui quale delle seguenti affermazioni è corretta.

Scegli un'alternativa:

- (●) a. L'analizzatore semantico effettua i controlli dei vincoli contestuali in un programma.
- ( ) b. L'analizzatore sintattico riconosce i token in un programma.
- ( ) c. L'analizzatore sintattico controlla che le variabili utilizzate in un programma siano state preventivamente dichiarate.
- ( ) d. L'analizzatore lessicale controlla la correttezza sintattica di un programma.

(PUNTI 2)

5) Sia data la seguente grammatica G=(X,V,S,P)

X = {a, b}   V = {S, A, B, C}

P = {S-->aS|bAS|aSb|aSc|B

A-->cB

B-->bB|bC

C-->a|aC}

Supponendo di avere un albero di derivazione T_w per una stringa w, e supponendo che l'altezza di T_w sia pari a 5, indicare la lunghezza massima di w, giustificando formalmente la risposta.

Per individuare la lunghezza massima della parola *w* è sufficiente richiamare il Lemma 4.1:

Sia G = (X, V, S, P) una grammatica C.F. e supponiamo che:

$$m=\max\{ |v| \mid A\text{->}v\in P\}$$

Sia T_w un albero di derivazione per una stringa *w* di L(G). Se l'altezza di T_w è al più uguale ad un intero *j*, allora: $|w| \leq m^j$.

Nel caso della G data si ha che la massima lunghezza delle parti destre delle produzioni è pari a 3, dunque m=3; l'altezza di T_w è 5, dunque possiamo concludere:

$$|w| \leq 3^5$$

(PUNTI 6)

6) Sia dato il linguaggio L1 su X = {a}

L1 = {a, aa}

e il linguaggio L2 su X = {a, b}

$$L2 = \{ab^n \mid n > 0\}$$

Determinare una espressione regolare che denota il linguaggio $L3=L1^* \cdot L2$

L1={a, aa} può essere denotato dall'espressione regolare R1=(a+aa) (S(R1)=L1).

L1*={a, aa}*={λ, a, aa, aaa,…}={a}* può essere denotato dall'espressione regolare R'1=a* (S(R'1)= L1*).

$L2=\{ab^n \mid n > 0\}=\{a\}\{b\}^+$ può essere denotato dall'espressione regolare R2=abb* (S(R2)=L2).

Dunque $L3=L1^*L2=\{a\}^*\{a\}\{b\}^+$ può essere denotato dall'espressione regolare R3= a*abb* (S(R3)=L3).

(PUNTI 4)

7) Fornire la definizione di albero di derivazione e illustrare il principio di sostituzione di sottoalberi nelle grammatiche context-free.

Sia $G = (X, V, S, P)$ una grammatica C.F. e $w \in X^*$ una stringa derivabile da S in G.

Dicesi *albero di derivazione* per w l'albero T_w avente le seguenti proprietà:

(1) la radice è etichettata con il simbolo iniziale S;

(2) ogni nodo interno (nodo non foglia) è etichettato con un simbolo di V (un nonterminale);

(3) ogni nodo foglia è etichettato con un simbolo di X (un terminale) o con λ;

(4) se un nodo N è etichettato con A, ed N ha k discendenti diretti $N_1, N_2,\ldots N_k$, etichettati con i simboli $A_1, A_2,\ldots A_k$ rispettivamente, allora la produzione $A \rightarrow A_1A_2\ldots A_k$ deve appartenere a P;

(5) la *stringa* w può essere ottenuta leggendo (e concatenando) le foglie dell'albero da sinistra a destra.

Sia $G = (X, V, S, P)$ una grammatica C.F. e $w \in X^*$ una stringa derivabile da S in G. Se denotiamo con T_w un albero di derivazione per w, il *principio di sostituzione di sottoalberi* nelle grammatiche context-free afferma che la sostituzione di un sottoalbero di T_w con un altro sottoalbero di T_w avente la stessa radice - più precisamente, i cui nodi radice siano etichettati con lo stesso simbolo nonterminale - produce un albero di derivazione valido per la grammatica G.

(PUNTI 5)